Using Clinical Characteristics to Identify Which Patients With Major Depressive Disorder Have a Higher Genetic Load for Three Psychiatric Disorders

Supplemental Information

NESDA & RADIANT-UK

Supplemental Text

Genotyping and genetic relationship matrix

Blood sample collection and DNA extraction methods have been previously described (1). Autosomal SNPs were genotyped on the Affymetrix 6.0 Human Genome-Wide SNP Array in three separate batches. Main quality control (QC) steps have been described previously (2; 3). Samples were excluded based on the following criteria: Affymetrix contrast QC <0.4; fell outside of the main cluster of a principal component reflecting a batch effect (3); missing rate > 5%; excess genome-wide heterozygosity or inbreeding levels (F<-0.10 or > 0.10); genotypes inconsistencies with reported gender; non-European/non-Dutch ancestry as indicated by principal component analysis. SNPs were excluded for the following reasons: probes mapped badly against NCBI Build 37/UCSC hg19; minor allele frequency <1%; missing rate > 5%; deviation from Hardy-Weinberg equilibrium (p<1e-5). Primary analyses included 497,347 SNPS. Additional stringent QC (as described in Milaneschi *et al.* (4)) was performed to build a genetic-relationship-matrix (GRM) in order to reduce the possibility that estimates from GRM-based analyses could be inflated by artifacts. The remaining 435,579 SNPs were used to build the GRM using GCTAv.1.24.1 (5).

Genomic profile risk scores (GPRS)

As previously described (4) results from the PGC were used to derive GPRS for MDD (6) (~8K cases, ~8K controls), bipolar disorder (7) (BIP; ~7K cases, ~9K controls) and schizophrenia (8) (SCZ2; ~36K cases, ~113K controls). Since NESDA and NTR samples contributed to MDD discovery GWAS, leave-one-out meta-analysis was performed with the Dutch GWAS cohort (9) excluded in order to remove any chance of overlap between discovery and target MDD samples. Risk scores (GPRS-thresholds) were obtained

following the method described by Purcell and colleagues (10). From the meta-analysis, SNPs with INFO >0.8 were retained. Independent SNPs among those overlapping between target and discovery samples were selected using p-value informed LD clumping (500 kb window, r^2=0.25) using PLINKv1.07 (10). For SCZ2 we included only one of the SNPs associated with schizophrenia from the extended major histocompatibility locus (eMCH, chr6:25-34Mb) given its complex LD structure. Eight sets of scores alleles were selected based on significance thresholds (Pt <.0001, <.001, <.005, <.01, <.05, <.1, <.5, <1) of the discovery samples associations. GPRS were calculated as the number of scores alleles weighted by effect sizes (log-OR) from the discovery statistics using PLINKv1.07. The number of SNPs included for each P-value threshold can be found in Table S2. The traditional method to build GPRS that is based on LD pruning and P thresholding may be limited in its predicting value by not taking into account information on LD structure (11). In addition, we therefore used LDpred a new approach to calculated genetic risk which is suggested to increase predictive accuracy (11). LDpred infers the posterior mean causal effect size of each marker using a prior on effect sizes and LD information from an external reference panel (1000 Genomes CEU in our analyses). Other parameters were set as follows: LD-radius=165 SNPs; Fraction of causal SNPs (P)=1 (infinitesimal model). Both GPRS thresholds and LDpred were standardized to a mean of zero and standard deviation of one to aid interpretation of results.

Statistical analyses

Firstly, focusing on MDD cases (*n*=1539) we regressed genetic risk (GPRS-thresholds and LDpred) over clinical characteristics of MDD (age at onset, duration of symptoms, family history, number of DSM symptoms, severity of symptoms, recurring episodes, stages) using linear regression analyses. In order to discard spurious correlations the following strategy was applied. First the GPRS-characteristic associations were tested computing permutation based empirical p-values: the characteristic values were permutated 10,000 times and the association with all GPRS re-estimated; empirical p-values were calculated as the number of simulations with p-values smaller than the original one divided by the number of permutations times the number of discovery traits times the number of polygenic scores

(10,000*3*9=270,000). Then, we calculated the false discovery rate (FDR) q values for each specific GPRS across the 6 main clinical characteristics (age at onset, duration, family history, number of DSM symptoms, severity of symptoms, recurring episodes). Only the GPRS-characteristic pairs showing the most consistent (higher number of significant tests across GPRS) profile of associations were selected for further analyses.

Replication sample

Imputed (HapMap3) genotype data of RADIANT-UK were processed according to QC steps described in detail in a previous publication by our group (12). A set of 754,464 QC-positive (MAF>0.01, missingness <2.0%, HWE P-value>1e-6 and imputation r^2>0.6) SNPS was pruned to a total of 76,201 independent SNPs with PLINK (--indep-pairwise 200 5 0.25) on which GPRS-SCZ were subsequently prepared. The number of SNPs included for each P-value threshold can be found in Table S6.

Sensitivity analyses: selection criteria of healthy controls in main analyses

Our findings highlighted the importance of selection of MDD cases in determining a stronger association with GPRS. It might be argued that, similarly to a more refined phenotypic definition of MDD cases, also the selection criteria applied to the control group may have an impact on the association with GPRS. We repeated the analyses in Table 1, and Figure 1, after including in the controls group 590 additional NTR participants selected using less strict criteria (scoring high on anxious depression factor score at one or multiple assessments, while for previous analyses only controls with low factor score on all assessments were selected; see MDD diagnoses section). Result patterns were substantially unchanged (Figure S3 shows the results for the proportion of explained liability variance) suggesting a minor impact on results of the controls selection criteria.

Clinical staging

Introduction

In addition to single MDD phenotype characteristics that indicate more homogenous groups of MDD cases, clinical staging combines different MDD phenotype characteristics (for example, duration and recurring (no/yes)) to create multiple homogenous groups of MDD organized in ascending stages of MDD progression (13–15). Previous studies using the staging model suggest that clinical staging has the power to detect differences across MDD patients categorized in phenotypically more homogenous subgroups, at least between those at-risk for MDD and full-threshold MDD cases (15–20); i.a. neuropsychological, brain structural (17; 18), and course trajectory differences (19; 20). To our knowledge it has not been examined whether clinical stages differ in their genetic load.

Methods

Finally, we applied a clinical staging algorithm, combining different clinical characteristics, e.g. recurring of MDD episodes, and chronic duration of MDD, to create multiple homogenous groups of MDD cases. Stages of MDD were based on the clinical staging model developed by McGorry *et al.* (15; 14; 21). Details on stage-assignment can be found in Figure S1. In short, cases were assigned to the three stages of MDD (2,3,4). Stage 2 (*n*=303) first MDD episode. Stage 3 (*n*=631) recurrent or relapse MDD episode. Stage 4 (*n*=605) chronic MDD, defined as an episode lasting longer than 2 years as indicated by the CIDI at baseline, or the life-chart during follow-up.

Discussion

Finally, there was no significant difference in genetic load across the staging model of MDD either. The staging model reflects progression of MDD by dividing cases into those with a first episode, recurrent episode and episodes with chronic duration (first or recurrent episode continuously present for 2 years) (19). Whether a more progressed stage of MDD defined according a clinical staging model is associated with increased genetic load has, to our knowledge, never been examined before. Based on our findings we suggest that stages of full-threshold MDD might be better defined on age at onset, number of DSM

symptoms, and/or severity of symptoms. In the bipolar-field (22), stages are proposed to be separated from each other based upon the presence/absence of symptoms between episodes (a similar measure to our average severity score).

NESDA

Supplemental Tables & Figures

Table S1. Associations between clinical characteristics of MDD and GPRS scores, in those who ever experienced MDD in life (*n*=1539)

	Age of Onset			Duration, mean over 10 yrs			Family History			Number of DSM symptoms (highest ever)			Severity of symptoms (average IDS) score			Recurring MDD (no/yes)			Stage of MDD (stage 2, stage 3, stage 4)		
GPRS MDD	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*
Pt < 0.0001	-.062	.031	**0.045**	.018	.026	0.476	-.035	.026	0.176	.016	.026	0.527	.044	.026	0.089	-.029	.026	0.253	-.019	.026	0.460
Pt < 0.001	-.002	.031	0.938	.013	.026	0.598	-.036	.026	0.162	.001	.026	0.962	.027	.026	0.300	.005	.026	0.852	-.012	.026	0.632
Pt < 0.005	-.040	.031	0.195	.039	.026	0.131	.005	.026	0.845	.010	.026	0.693	.037	.026	0.154	.044	.026	0.085	.016	.026	0.531
Pt < 0.01	-.034	.031	0.277	.063	.026	**0.014**	.009	.026	0.741	.044	.026	0.090	.059	.026	**0.022**	.012	.026	0.653	.025	.026	0.328
Pt < 0.05	-.008	.031	0.798	.010	.026	0.697	.041	.026	0.112	.063	.026	**0.014**	.041	.026	0.108	.026	.026	0.315	.027	.026	0.291
Pt < 0.1	-.005	.031	0.871	.020	.026	0.441	.014	.026	0.576	.049	.026	0.060	.039	.026	0.123	.031	.025	0.221	.031	.026	0.222
Pt < 0.5	-.018	.031	0.558	.016	.025	0.528	.005	.025	0.844	.058	.026	**0.025**	.044	.025	0.084	.029	.025	0.246	.026	.025	0.306
Pt < 1.0	-.030	.031	0.324	.021	.025	0.412	.004	.025	0.890	.062	.026	**0.015**	.047	.025	0.063	.035	.025	0.162	.030	.025	0.231
LDpred	-.029	.031	0.346	.007	.026	0.772	.003	.026	0.918	.063	.026	**0.015**	.056	.025	**0.029**	.018	.025	0.489	.006	.026	0.819
GPRS BIP	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*
Pt < 0.0001	-.027	.031	0.385	-.020	.026	0.442	.025	.026	0.329	-.004	.026	0.883	-.033	.026	0.199	.024	.026	0.350	-.031	.026	0.230
Pt < 0.001	-.072	.031	**0.020**	-.008	.026	0.763	.011	.026	0.668	.039	.026	0.138	-.020	.026	0.437	.068	.026	**0.008**	.015	.026	0.547
Pt < 0.005	-.115	.031	**$2e^{-04}$**	.001	.026	0.964	-.010	.026	0.704	.038	.026	0.147	-.002	.026	0.925	.061	.026	**0.017**	.027	.026	0.292
Pt < 0.01	-.088	.031	**0.005**	-.020	.026	0.435	.003	.026	0.896	.023	.026	0.366	-.018	.026	0.487	.040	.026	0.116	-.004	.026	0.886
Pt < 0.05	-.064	.031	**0.040**	.021	.026	0.416	.028	.026	0.273	.065	.026	**0.012**	.001	.026	0.961	.041	.026	0.109	.014	.026	0.590
Pt < 0.1	-.051	.031	0.102	.016	.026	0.543	.027	.026	0.302	.058	.026	**0.024**	-.009	.026	0.717	.031	.026	0.222	.008	.026	0.742
Pt < 0.5	-.046	.031	0.141	.030	.026	0.240	-.001	.026	0.957	.041	.026	0.111	-.006	.026	0.807	.032	.026	0.206	.000	.026	0.987
Pt < 1.0	-.043	.031	0.169	.032	.026	0.211	.000	.026	0.988	.038	.026	0.149	-.007	.026	0.784	.038	.026	0.135	.003	.026	0.895
LDpred	-.062	.031	**0.045**	.033	.026	0.193	.013	.026	0.612	.054	.026	**0.037**	-.008	.026	0.762	.049	.026	0.055	.017	.026	0.499
GPRS SCZ	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*	Beta	SE	P-val*
Pt < 0.0001	-.048	.031	0.122	.045	.026	0.079	.034	.026	0.181	.038	.026	0.145	.062	.026	**0.014**	-.005	.026	0.850	.046	.026	0.073
Pt < 0.001	-.050	.031	0.108	.046	.026	0.075	.028	.026	0.285	.047	.026	0.071	.067	.026	**0.009**	-.022	.026	0.398	.025	.026	0.329
Pt < 0.005	-.030	.031	0.334	.043	.026	0.096	.019	.026	0.449	.050	.026	0.053	.064	.026	**0.012**	-.034	.026	0.183	.006	.026	0.800
Pt < 0.01	-.032	.031	0.300	.048	.026	0.062	.015	.026	0.558	.058	.026	**0.025**	.089	.025	**$3.74e^{-4}$**	-.043	.025	0.088	.004	.026	0.879
Pt < 0.05	-.033	.031	0.290	.022	.026	0.380	-.002	.026	0.949	.056	.026	**0.030**	.050	.025	**0.050**	-.046	.025	0.067	-.006	.026	0.797
Pt < 0.1	-.022	.031	0.478	.010	.025	0.689	.004	.025	0.891	.041	.026	0.110	.028	.025	0.261	-.056	.025	**0.025**	-.022	.025	0.394
Pt < 0.5	-.019	.031	0.526	.008	.025	0.755	-.007	.025	0.798	.044	.026	0.088	.016	.025	0.534	-.044	.025	0.082	-.005	.025	0.843
Pt < 1.0	-.021	.031	0.488	.007	.025	0.770	-.005	.025	0.845	.044	.026	0.088	.015	.025	0.552	-.042	.025	0.093	-.007	.025	0.791
LDpred	-.017	.031	0.586	.021	.025	0.418	-.003	.025	0.910	.064	.026	**0.013**	.031	.025	0.216	-.028	.025	0.265	.018	.025	0.471

Table S1

BIP= Bipolar disorder; MDD= Major Depressive Disorder; *n*= number; GPRS= Genomic Profile Risk Score; SCZ= Schizophrenia; SE= Standardized Error.

Analyses are adjusted for Year of Birth, Gender; and three Principal Components.

P-val*= Empirical P-value, P-value derived after 270,000 permutations (9 Pt Thresholds * 3 Traits * 10,000 permutations).

Additional FDR-adjusted analyses (adjusted for 6 characteristics) showed that with a significance level of <0.10 for the link:

GPRS-MDD – Number of DSM Symptoms 3 of the 4 remained significant, namely Pt-thresholds <.05, <1.0, LDpred

GPRS-BIP – Age at Onset 3 of the 5 remained significant, namely Pt-thresholds <.001, <.005, <.01

GPRS-BIP – Number of DSM Symptoms 1 of 3 remained significant, namely Pt-thresholds <.05

GPRS-SCZ – Severity of Symptoms (mean IDS-score) 4 of 5 remained significant, namely Pt-thresholds <.0001, <.001, <.005, <.01

GPRS-SCZ – Number of DSM Symptoms 2 of 3 remained significant, namely Pt-thresholds <.01, LDpred

Table S2. Number of SNPs included in the GPRS according to eight significance thresholds of the discovery samples association

	NESDA		
Pt threshold	MDD	BIP	SCZ
Pt <.0001	37	79	807
Pt <.001	286	479	2229
Pt <.005	1258	1654	5196
Pt <.01	2296	2911	7659
Pt <.05	9854	10974	20344
Pt <.1	18160	19154	31776
Pt <.5	71516	68565	91636
Pt <1 (all SNPs)	112018	104488	134114

Table S3. Association between genetic risk scores (GPRS) and MDD subgroups stratified according to clinical characteristics in NESDA

GPRS MDD

P-value threshold	<0.0001 OR (CI)	P-value	<0.001 OR (CI)	P-value	<0.005 OR (CI)	P-value	<0.01 OR (CI)	P-value	<0.05 OR (CI)	P-value	<0.1 OR (CI)	P-value	<0.5 OR (CI)	P-value	<1.0 OR (CI)	P-value	LDpred OR (CI)	P-value
MDD all cases (n=1539)	1.10 (1.02-1.18)	1.03E-02	1.02 (0.95-1.09)	6.22E-01	1.08 (1.01-1.16)	2.99E-02	1.09 (1.02-1.17)	1.32E-02	1.18 (1.10-1.26)	6.75E-06	1.20 (1.12-1.29)	6.24E-07	1.21 (1.13-1.30)	1.93E-07	1.21 (1.13-1.31)	1.20E-07	1.24 (1.15-1.33)	4.19E-09
DSM 5/6 (n=244)	1.04 (0.91-1.19)	5.95E-01	0.99 (0.86-1.13)	8.46E-01	1.00 (0.87-1.14)	9.96E-01	0.95 (0.83-1.08)	4.10E-01	0.99 (0.87-1.14)	9.08E-01	1.05 (0.91-1.20)	5.03E-01	1.08 (0.94-1.23)	2.80E-01	1.07 (0.93-1.22)	3.57E-01	1.11 (0.97-1.27)	1.40E-01
DSM 7 (n=302)	1.12 (0.99-1.27)	6.30E-02	1.05 (0.93-1.19)	4.07E-01	1.17 (1.03-1.32)	1.42E-02	1.18 (1.04-1.33)	9.03E-03	1.22 (1.08-1.38)	1.25E-03	1.26 (1.11-1.42)	2.45E-04	1.23 (1.09-1.39)	8.04E-04	1.24 (1.10-1.41)	4.97E-04	1.22 (1.08-1.38)	1.32E-03
DSM 8 (n=442)	1.12 (1.01-1.25)	3.26E-02	1.02 (0.91-1.13)	7.52E-01	1.09 (0.98-1.21)	1.11E-01	1.10 (0.99-1.22)	8.05E-02	1.17 (1.06-1.30)	3.16E-03	1.18 (1.06-1.31)	2.79E-03	1.18 (1.06-1.31)	2.70E-03	1.18 (1.06-1.31)	1.96E-03	1.20 (1.08-1.34)	7.19E-04
DSM 9 (n=499)	1.09 (0.99-1.21)	8.25E-02	1.01 (0.91-1.12)	8.69E-01	1.07 (0.97-1.18)	1.97E-01	1.12 (1.02-1.24)	2.30E-02	1.25 (1.13-1.39)	1.29E-05	1.26 (1.14-1.40)	5.38E-06	1.30 (1.18-1.44)	3.15E-07	1.31 (1.18-1.45)	1.71E-07	1.36 (1.22-1.50)	4.23E-09
DSM-high (n=941)	1.11 (1.02-1.20)	1.11E-02	1.02 (0.94-1.10)	6.71E-01	1.08 (1.00-1.17)	6.46E-02	1.11 (1.02-1.21)	1.11E-02	1.22 (1.12-1.33)	3.03E-06	1.22 (1.12-1.33)	2.36E-06	1.24 (1.14-1.35)	4.72E-07	1.25 (1.15-1.35)	2.35E-07	1.29 (1.18-1.40)	4.04E-09

GPRS BIP

P-value threshold	<0.0001 OR (CI)	P-value	<0.001 OR (CI)	P-value	<0.005 OR (CI)	P-value	<0.01 OR (CI)	P-value	<0.05 OR (CI)	P-value	<0.1 OR (CI)	P-value	<0.5 OR (CI)	P-value	<1.0 OR (CI)	P-value	LDpred OR (CI)	P-value
MDD all cases (n=1539)	1.04 (0.97-1.12)	2.53E-01	1.03 (0.96-1.10)	4.54E-01	1.05 (0.98-1.13)	1.47E-01	1.08 (1.01-1.16)	3.18E-02	1.09 (1.02-1.17)	1.67E-02	1.11 (1.03-1.19)	3.92E-03	1.13 (1.05-1.21)	7.68E-04	1.13 (1.06-1.22)	4.80E-04	1.16 (1.08-1.25)	2.97E-05
DSM 5/6 (n=244)	1.11 (0.97-1.27)	1.35E-01	0.97 (0.85-1.11)	6.40E-01	1.01 (0.88-1.16)	8.53E-01	1.05 (0.91-1.20)	4.98E-01	1.01 (0.88-1.16)	8.50E-01	1.07 (0.94-1.23)	3.01E-01	1.11 (0.97-1.27)	1.45E-01	1.11 (0.97-1.27)	1.39E-01	1.08 (0.94-1.24)	2.76E-01
DSM 7 (n=302)	1.00 (0.88-1.13)	9.89E-01	0.98 (0.87-1.11)	8.05E-01	0.95 (0.84-1.08)	4.47E-01	1.00 (0.88-1.13)	9.75E-01	0.98 (0.86-1.11)	7.17E-01	0.98 (0.87-1.11)	7.23E-01	1.02 (0.90-1.15)	7.63E-01	1.03 (0.91-1.17)	6.12E-01	1.05 (0.93-1.19)	4.22E-01
DSM 8 (n=442)	0.98 (0.89-1.09)	7.68E-01	1.03 (0.92-1.14)	6.19E-01	1.12 (1.01-1.25)	3.08E-02	1.14 (1.03-1.27)	1.42E-02	1.13 (1.02-1.26)	2.17E-02	1.13 (1.01-1.25)	2.80E-02	1.17 (1.05-1.30)	3.93E-03	1.18 (1.06-1.31)	2.21E-03	1.23 (1.11-1.37)	1.50E-04
DSM 9 (n=499)	1.09 (0.98-1.20)	1.05E-01	1.08 (0.98-1.20)	1.14E-01	1.09 (0.99-1.21)	8.78E-02	1.11 (1.00-1.22)	5.10E-02	1.20 (1.08-1.33)	4.30E-04	1.23 (1.11-1.36)	6.85E-05	1.20 (1.08-1.32)	5.16E-04	1.19 (1.08-1.32)	7.41E-04	1.23 (1.11-1.36)	6.00E-05
DSM-high (n=941)	1.04 (0.96-1.13)	3.70E-01	1.06 (0.98-1.15)	1.37E-01	1.11 (1.03-1.21)	1.08E-02	1.12 (1.04-1.22)	5.08E-03	1.17 (1.08-1.27)	1.67E-04	1.18 (1.09-1.29)	5.12E-05	1.19 (1.10-1.29)	3.05E-05	1.19 (1.10-1.29)	2.62E-05	1.23 (1.13-1.34)	7.24E-07
AaO >37yrs (n=392)	1.00 (0.89-1.13)	9.95E-01	0.93 (0.83-1.05)	2.62E-01	0.92 (0.82-1.04)	1.99E-01	0.95 (0.85-1.07)	4.29E-01	1.03 (0.91-1.16)	6.27E-01	1.07 (0.95-1.21)	2.43E-01	1.11 (0.98-1.25)	9.14E-02	1.11 (0.99-1.25)	8.45E-02	1.09 (0.97-1.24)	1.43E-01
AaO 26-37yrs (n=380)	0.99 (0.88-1.10)	8.17E-01	0.99 (0.89-1.11)	9.00E-01	1.02 (0.92-1.15)	6.72E-01	1.08 (0.97-1.21)	1.73E-01	1.05 (0.94-1.18)	3.66E-01	1.04 (0.93-1.17)	4.77E-01	1.09 (0.98-1.22)	1.22E-01	1.10 (0.98-1.23)	9.14E-02	1.13 (1.01-1.26)	3.56E-02
AaO 18-25yrs (n=398)	1.05 (0.94-1.18)	3.46E-01	1.05 (0.94-1.17)	3.79E-01	1.03 (0.92-1.15)	5.91E-01	1.04 (0.93-1.16)	5.00E-01	1.04 (0.93-1.16)	5.01E-01	1.07 (0.96-1.20)	2.21E-01	1.11 (0.99-1.24)	6.45E-02	1.12 (1.01-1.25)	3.98E-02	1.15 (1.03-1.29)	1.14E-02
AaO <18yrs (n=360)	1.11 (0.99-1.25)	6.87E-02	1.11 (0.99-1.25)	6.39E-02	1.25 (1.11-1.40)	1.96E-04	1.25 (1.11-1.40)	1.76E-04	1.25 (1.11-1.40)	1.49E-04	1.27 (1.13-1.43)	4.52E-05	1.22 (1.09-1.37)	7.73E-04	1.21 (1.08-1.36)	9.39E-04	1.29 (1.15-1.45)	1.30E-05
AaO-young (n=758)	1.08 (0.99-1.18)	7.26E-02	1.09 (1.00-1.19)	6.09E-02	1.13 (1.04-1.24)	5.59E-03	1.14 (1.04-1.24)	3.83E-03	1.13 (1.04-1.24)	5.36E-03	1.16 (1.06-1.27)	8.38E-04	1.16 (1.06-1.27)	7.48E-04	1.17 (1.07-1.27)	5.31E-04	1.22 (1.12-1.33)	7.85E-06

GPRS SCZ

P-value threshold	<0.0001 OR (CI)	P-value	<0.001 OR (CI)	P-value	<0.005 OR (CI)	P-value	<0.01 OR (CI)	P-value	<0.05 OR (CI)	P-value	<0.1 OR (CI)	P-value	<0.5 OR (CI)	P-value	<1.0 OR (CI)	P-value	LDpred OR (CI)	P-value
MDD all cases (n=1539)	1.18 (1.10-1.27)	5.76E-06	1.22 (1.14-1.31)	3.78E-08	1.28 (1.19-1.38)	1.82E-11	1.28 (1.19-1.38)	2.27E-11	1.28 (1.19-1.38)	2.27E-11	1.31 (1.21-1.41)	7.18E-13	1.31 (1.22-1.41)	3.37E-13	1.31 (1.22-1.41)	6.89E-13	1.35 (1.26-1.46)	6.15E-16
DSM 5/6 (n=244)	1.16 (1.01-1.33)	3.71E-02	1.17 (1.02-1.35)	2.21E-02	1.15 (1.00-1.32)	4.51E-02	1.13 (0.98-1.29)	8.78E-02	1.16 (1.01-1.33)	3.35E-02	1.21 (1.06-1.39)	5.01E-03	1.22 (1.06-1.39)	4.81E-03	1.21 (1.05-1.38)	6.58E-03	1.21 (1.05-1.39)	6.78E-03
DSM 7 (n=302)	1.10 (0.97-1.24)	1.32E-01	1.13 (1.00-1.28)	4.68E-02	1.26 (1.11-1.43)	2.93E-04	1.26 (1.11-1.42)	3.45E-04	1.22 (1.08-1.38)	1.72E-03	1.25 (1.10-1.41)	4.56E-04	1.28 (1.13-1.44)	1.18E-04	1.28 (1.13-1.45)	1.01E-04	1.27 (1.12-1.44)	1.60E-04
DSM 8 (n=442)	1.18 (1.06-1.32)	1.97E-03	1.22 (1.09-1.35)	3.42E-04	1.31 (1.17-1.46)	1.08E-06	1.31 (1.18-1.46)	6.25E-07	1.31 (1.17-1.45)	1.04E-06	1.34 (1.20-1.49)	9.39E-08	1.33 (1.19-1.48)	2.49E-07	1.32 (1.19-1.47)	4.19E-07	1.37 (1.23-1.52)	1.13E-08
DSM 9 (n=499)	1.26 (1.14-1.40)	9.20E-06	1.33 (1.20-1.47)	4.39E-08	1.36 (1.23-1.51)	3.14E-09	1.37 (1.24-1.52)	1.32E-09	1.37 (1.24-1.52)	1.58E-09	1.36 (1.23-1.51)	2.62E-09	1.38 (1.25-1.53)	7.24E-10	1.37 (1.24-1.52)	1.61E-09	1.47 (1.33-1.63)	2.26E-13
DSM-high (n=941)	1.22 (1.13-1.33)	2.00E-06	1.28 (1.18-1.40)	4.27E-09	1.35 (1.24-1.46)	5.68E-12	1.35 (1.24-1.47)	2.77E-12	1.35 (1.24-1.47)	4.31E-12	1.35 (1.25-1.47)	1.94E-12	1.36 (1.25-1.49)	1.01E-12	1.36 (1.25-1.48)	2.38E-12	1.44 (1.32-1.57)	1.45E-16
IDS <13.00 (n=384)	1.10 (0.99-1.23)	8.36E-02	1.11 (0.99-1.24)	6.79E-02	1.16 (1.03-1.29)	1.18E-02	1.13 (1.01-1.26)	3.70E-02	1.20 (1.07-1.34)	1.51E-03	1.25 (1.12-1.40)	1.02E-04	1.26 (1.13-1.41)	6.18E-05	1.25 (1.12-1.40)	8.27E-05	1.26 (1.12-1.41)	6.78E-05
IDS >13.00 - ≤20.25 (n=385)	1.11 (0.99-1.24)	6.80E-02	1.19 (1.06-1.33)	2.67E-03	1.23 (1.10-1.38)	2.76E-04	1.21 (1.08-1.36)	7.76E-04	1.22 (1.09-1.36)	6.66E-04	1.27 (1.14-1.42)	2.71E-05	1.28 (1.15-1.44)	1.38E-05	1.29 (1.15-1.44)	1.24E-05	1.34 (1.19-1.50)	5.06E-07
IDS >20.25 - ≤29.00 (n=387)	1.25 (1.12-1.40)	1.16E-04	1.29 (1.15-1.44)	1.42E-05	1.36 (1.22-1.53)	1.06E-07	1.38 (1.23-1.54)	4.51E-08	1.39 (1.24-1.55)	1.85E-08	1.39 (1.24-1.56)	1.14E-08	1.43 (1.28-1.61)	6.16E-10	1.42 (1.27-1.59)	1.83E-09	1.46 (1.30-1.64)	7.62E-11
IDS >29.00 - 84 (n=377)	1.28 (1.14-1.43)	2.72E-05	1.34 (1.19-1.50)	6.95E-07	1.40 (1.25-1.57)	8.93E-09	1.45 (1.29-1.63)	2.88E-10	1.34 (1.20-1.50)	5.02E-07	1.32 (1.18-1.48)	1.69E-06	1.30 (1.16-1.45)	8.85E-06	1.29 (1.15-1.45)	1.10E-05	1.38 (1.23-1.55)	2.76E-08
IDS-high (n=764)	1.27 (1.16-1.39)	1.70E-07	1.32 (1.21-1.45)	8.80E-10	1.39 (1.27-1.52)	9.60E-13	1.42 (1.29-1.55)	5.95E-14	1.38 (1.26-1.51)	5.44E-12	1.36 (1.24-1.49)	2.14E-11	1.37 (1.25-1.50)	1.76E-11	1.36 (1.24-1.49)	4.57E-11	1.44 (1.31-1.58)	1.41E-14

Table S3: OR, 95%CI and P-values from binary (MDDall, DSM-high, IDS-high, AaO -young) and multinomial (subgroups) logistic regression (reference=controls, *n*=1792); adjusted for year of birth, gender and 3 principal components.

AaO= Age at Onset, AaO-young= MDD cases with a young age at onset (<26yrs); CI= Confidence Interval; DSM-high= MDD cases with a high number of DSM (8 or 9) symptoms; IDS-high= MDD cases with a high severity (IDS > 20.25) score; MDDall= all MDD cases; *n*=number; OR=Odds Ratio.
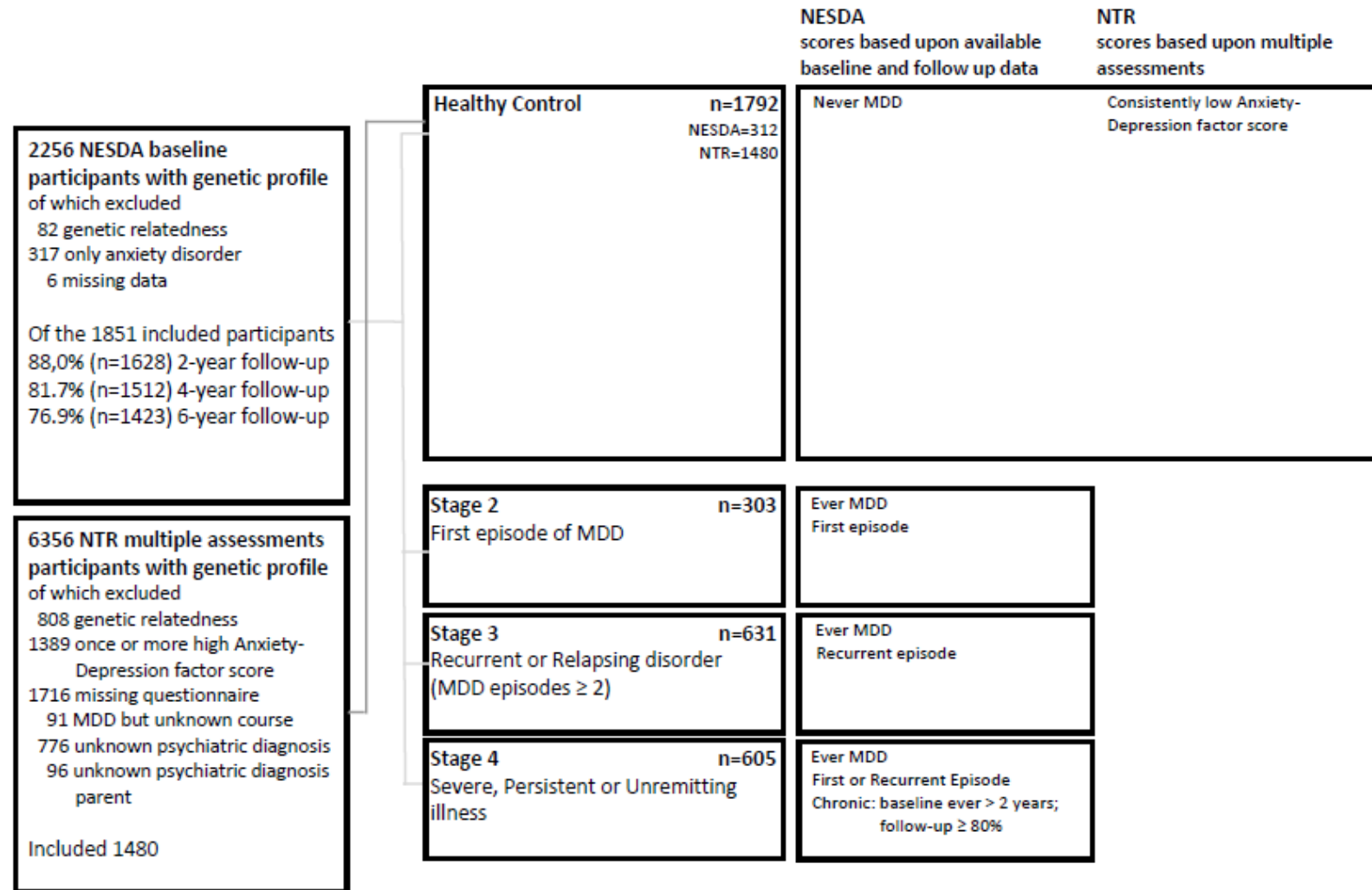
Figure S1. Assignment of NESDA and NTR participants to the clinical stages of MDD

MDD = Major Depressive Disorder; Chronic if MDD symptoms present at baseline ever more than 2 years, at follow-up if symptoms are present for ≥ 80% of follow-up time.
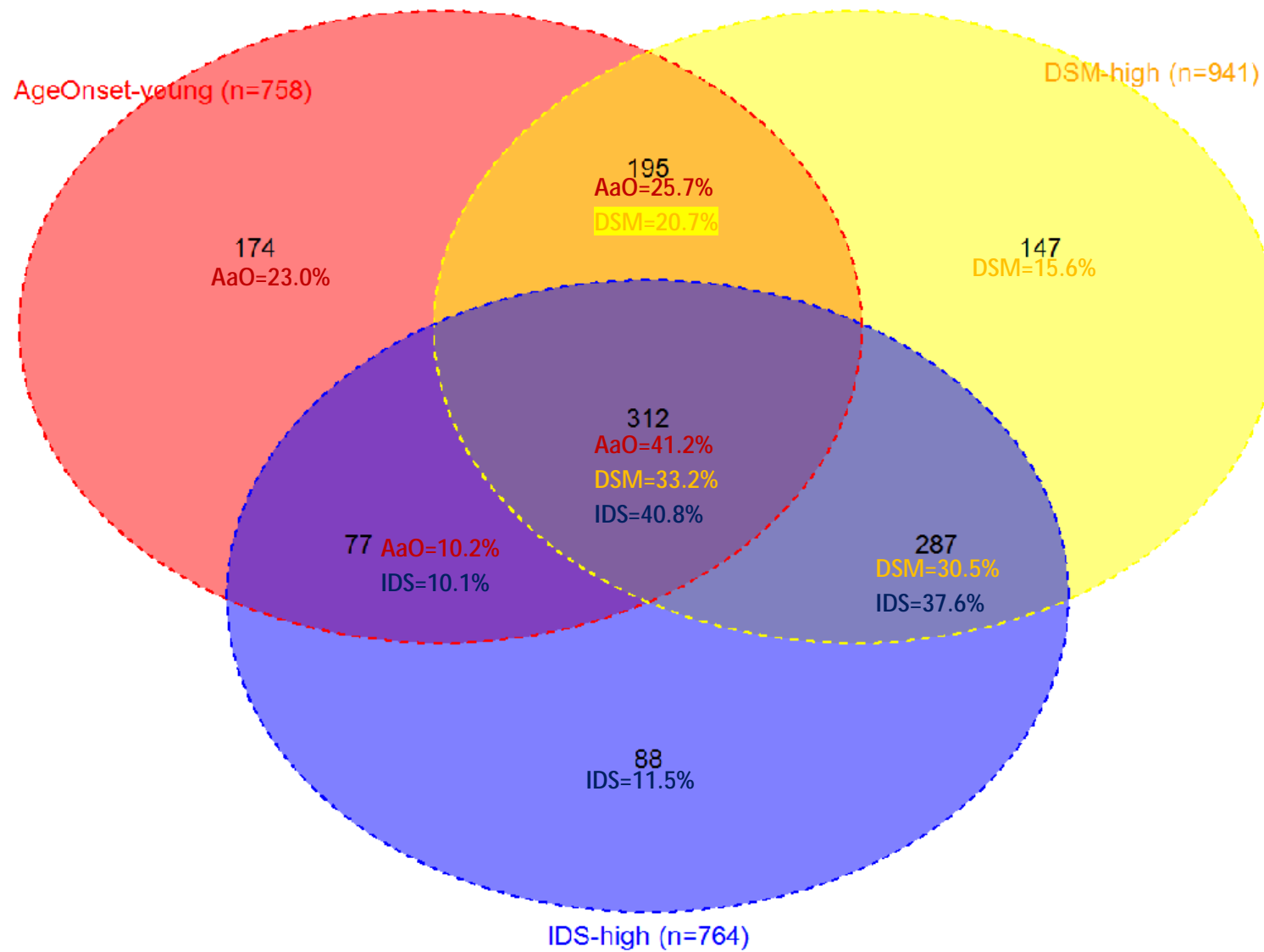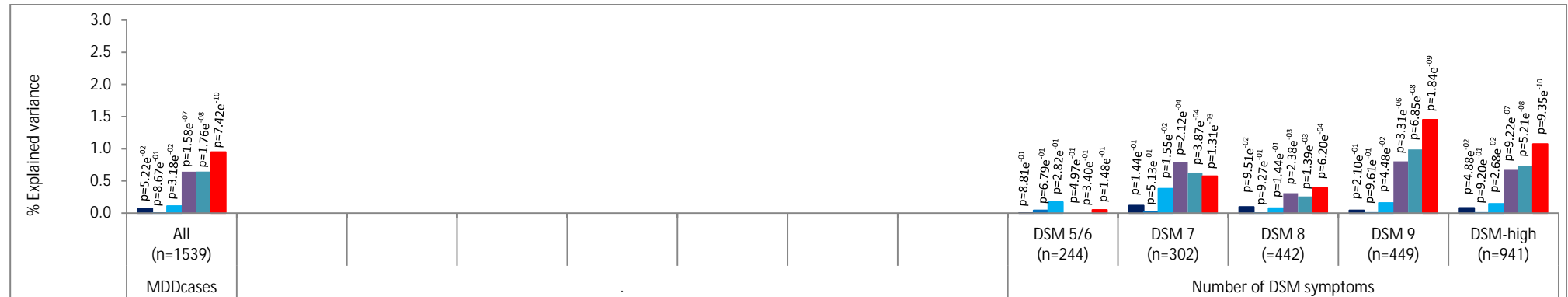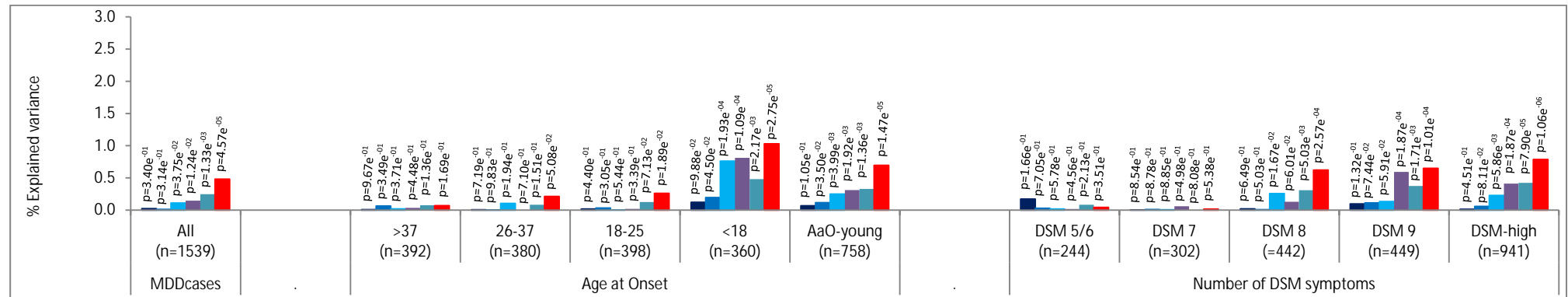
Figure S2. Overlap between subgroups of MDD cases selected according different clinical characteristics

GPRS-MDD

GPRS-BIP

GPRS-SCZ
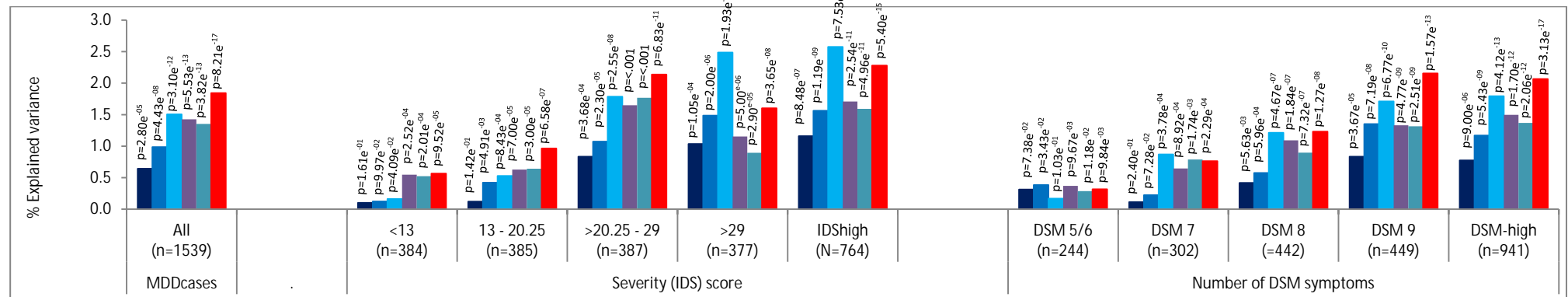
Legend: ■ Pt<0.0001 ■ Pt<0.001 ■ Pt<0.01 ■ Pt<0.1 ■ Pt<1.0 ■ LDpred

Figure S3. Sensitivity analyses: explained variance after adding extra (n=590) controls

Explained variance assuming a liability threshold model and K=0.18 (MDD), K=0.18/4 (DSM, IDS and AaO quartiles), K=0.18/2 (DSM-high, IDS-high, AaO-young)
P-values from binary (MDDall, DSM-high, IDS-high, AaO-young) and multinomial (subgroups) logistic regression (reference=controls, n=2382); adjusted for year of birth, gender and 3 principal components

Table S4. Genetic covariance between psychiatric traits and subgroups of MDD patients selected according to clinical characteristics

AVENGEME input in the estimatePolygenicModel formula
Schizophrenia Number of independent schizophrenia SNPs=134114, Size training (PGC) sample *n*=150064, 36989 cases, 113075 controls (8) Schizophrenia genetic variance = 0.23, h2-SNP (23) Schizophrenia prevalence= 0.01 (23)
Bipolar Disorder Number of independent bipolar SNPs=104488 Size training (PGC) sample *n*=16731, 7481 cases, 9250 controls (24) Bipolar genetic variance = 0.25, h2-SNP (23) Bipolar prevalence= 0.01 (23)
Input specific for DSM-high Size target (NESDA) sample *n*=2733, 941 cases, 1792 controls. DSM-high prevalence=0.12, derived by taking ~66% (percentage with DSM-high in NESDA) of the prevalence of MDD in the Netherlands K=0.18
Input specific for IDS-high Size target (NESDA) sample *n*=2556, 764 cases, 1792 controls. IDS-high prevalence=0.09, derived by taking ~50% (percentage with IDS-high in NESDA) of the prevalence of MDD in the Netherlands K=0.18
Input specific for AaO-young Size target (NESDA) sample *n*=2550, 758 cases, 1792 controls. AaO-young prevalence=0.09, derived by taking ~50% (percentage with AaO-young in NESDA) of the prevalence of MDD in the Netherlands K=0.18

Table S5. Calculation of Power of GPRS for Schizophrenia to detect an association with DSM-high in NESDA

AVENGEME input in the polygenescore formula
Schizophrenia
Number of independent schizophrenia SNPs=134114,
Size training (PGC) sample *n*=150064, 36989 cases, 113075 controls (8)
Schizophrenia genetic variance = 0.23, h2-SNP (23)
Schizophrenia prevalence= 0.01 (23)
MDD DSM-high
Size target (NESDA) sample *n*=2733, 941 cases, 1792 controls.
DSM-high prevalence=0.12, derived by taking ~66% (percentage with DSM-high in NESDA) of the prevalence of MDD in the Netherlands K=0.18
Estimated genetic covariance between Schizophrenia and MDD DSM-high=0.1205118

RADIANT-UK

Supplemental Tables & Figures

Table S6. Number of SNPs included in the GPRS according to eight significance thresholds of the discovery samples association

	RADIANT-UK
Pt threshold	SCZ1
Pt <.0001	315
Pt <.001	902
Pt <.005	2105
Pt <.01	3173
Pt <.05	8647
Pt <.1	13694
Pt <.5	43711
Pt <1 (all SNPs)	76201

Table S7. Association between genetic risk scores (GPRS) and MDD with high number of DSM endorsed symptoms in RADIANT-UK

GPRS SCZ																
P-value threshold	<0.0001		<0.001		<0.005		<0.01		<0.05		<0.1		<0.5		<1.0	
	OR (CI)	P-value	OR (CI)	P-value	OR (CI)	P-value	OR (CI)	P-value	OR (CI)	P-value	OR (CI)	P-value	OR (CI)	P-value	OR (CI)	P-value
MDD all cases (*n*=1602)	1.08 (0.99-1.17)	7.16E-02	1.14 (1.06-1.24)	9.18E-04	1.21 (1.12-1.31)	1.25E-06	1.22 (1.13-1.32)	3.90E-07	1.23 (1.14-1.33)	2.27E-07	1.25 (1.16-1.36)	1.45E-08	1.23 (1.14-1.34)	1.15E-07	1.22 (1.13-1.32)	3.26E-07
DSM-high (*n*=878)	1.09 (0.99-1.21)	6.55E-02	1.16 (1.06-1.28)	1.26E-03	1.25 (1.14-1.37)	1.90E-06	1.25 (1.14-1.38)	1.31E-06	1.27 (1.16-1.40)	2.81E-07	1.31 (1.19-1.44)	1.57E-08	1.27 (1.15-1.39)	6.35E-07	1.25 (1.14-1.37)	1.66E-06

Table S7: OR, 95%CI and P-values from binary (MDDall, DSM-high) logistic regression (reference=controls, *n*=1390); adjusted for year of birth, gender and 10 principal components.

CI= Confidence Interval; DSM-high= MDD cases with a high number of DSM (8 or 9) symptoms; MDDall= all MDD cases; *n*=number; OR= Odds Ratio.

Table S8. Calculation of Power of GPRS for Schizophrenia to detect an association with DSM-high in RADIANT-UK

AVENGEME input in the polygenescore formula
Schizophrenia
Number of independent schizophrenia SNPs=76201,
Size training (PGC) sample *n*=150064, 36989 cases, 113075 controls (8)
Schizophrenia genetic variance = 0.23, h2-SNP (23)
Schizophrenia prevalence= 0.01 (23)
MDD DSM-high
Size target (RADIANT-UK) sample *n*=2268, 878 cases, 1390 controls.
DSM-high prevalence=0.12, derived by taking ~60% (percentage with DSM-high in RADIANT-UK) of the prevalence of MDD in the PGC sample K=0.21
Estimated genetic covariance between Schizophrenia and MDD DSM-high=0.08871459, derived from estimatePolygenicModel formula by Dudbridge *et al.* (25)

POOLED-ANALYSES

Supplemental Tables & Figures

NESDA (*n*=1487)	RADIANT-UK (*n*=1462)

Mean: 7.75
Median: 8.00
1st Q: 7.0
3rd Q: 9.0

DSM
#5/#6: 244 = 16.4%
#7: 302 = 20.3%
#8: 442 = 29.7%
#9: 499 = 33.6%

DSM-high: 941 = 63.3%

Mean: 7.66
Median: 8.00
1st Q: 7.0
3rd Q: 9.0

DSM
#5/#6: 229 = 15.6%
#7: 355 = 24.3%
#8: 481 = 32.9%
#9: 397 = 27.2%
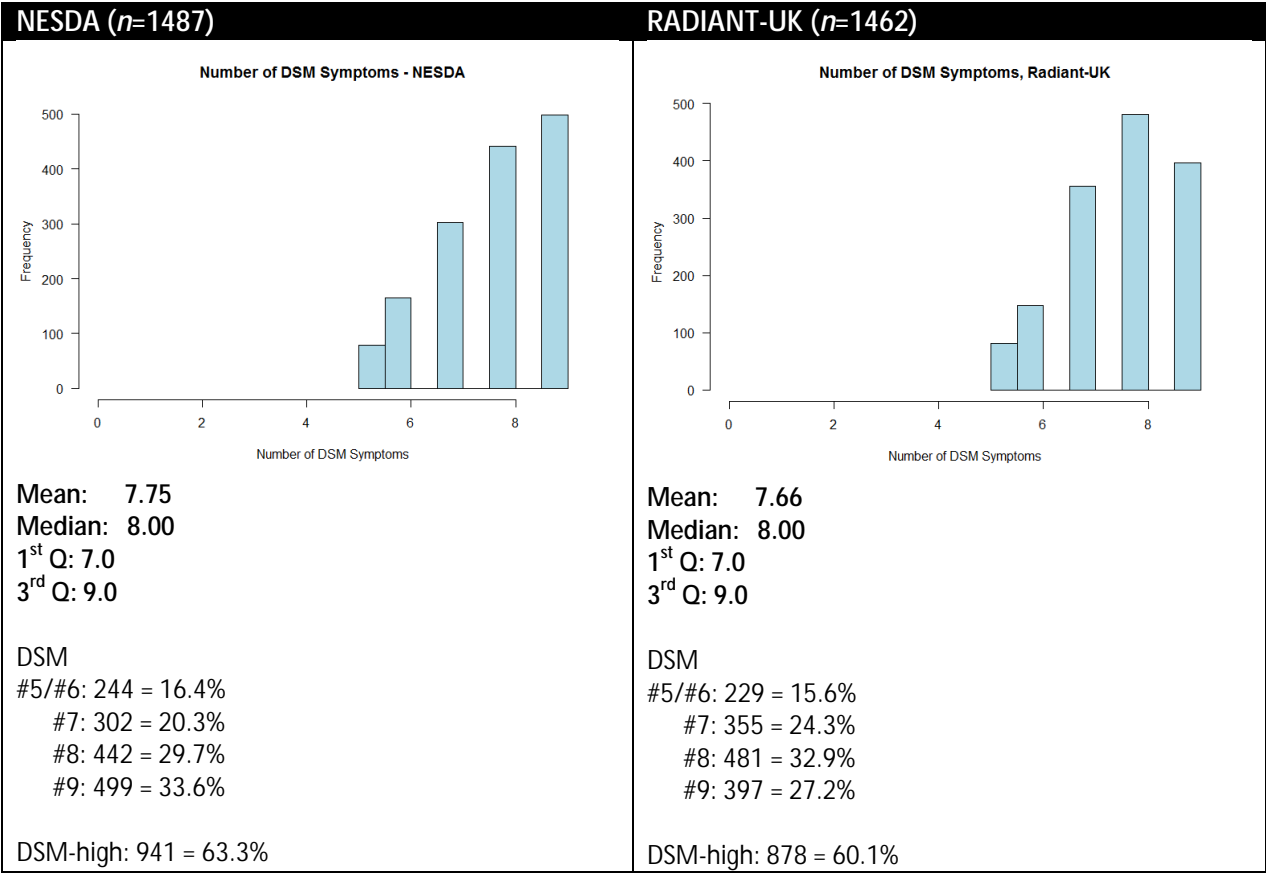
DSM-high: 878 = 60.1%

Figure S4. Distribution of number of DSM symptoms per study

Table S9. Pooled data analyses output, of the odds ratios derived from logistic regression analyses comparing the GPRS-SCZ in MDD overall vs. controls and DSM-high vs. controls both in NESDA and RADIANT-UK

	MDD all cases vs. controls			DSM-high symptoms (DSM ≥ 8) vs. controls		
	Odds Ratio	CI 95%	P-value	Odds Ratio	CI 95%	P-value
NESDA	Cases: 1539, Controls: 1792			Cases 941, Controls 1792		
Pt < 0.0001	1.18	1.10 - 1.27	5.76e-06	1.22	1.13 - 1.33	2.00e-06
Pt < 0.001	1.22	1.14 - 1.31	3.78e-08	1.28	1.18 - 1.40	4.27e-09
Pt < 0.005	1.28	1.19 - 1.38	1.82e-11	1.35	1.24 - 1.46	5.68e-12
Pt < 0.01	1.28	1.19 - 1.38	2.27e-11	1.35	1.24 - 1.47	2.77e-12
Pt < 0.05	1.28	1.19 - 1.38	2.27e-11	1.35	1.24 - 1.47	4.31e-12
Pt < 0.1	1.31	1.21 - 1.41	7.18e-13	1.35	1.25 - 1.47	1.94e-12
Pt < 0.5	1.31	1.22 - 1.41	3.37e-13	1.36	1.25 - 1.49	1.01e-12
Pt < 1.0	1.31	1.22 - 1.41	6.89e-13	1.36	1.25 - 1.48	2.38e-12
RADIANT-UK	Cases: 1602, Controls 1390			Cases: 878, Controls 1390		
Pt < 0.0001	1.08	0.99 - 1.17	7.16e-02	1.09	0.99 - 1.21	6.55e-02
Pt < 0.001	1.14	1.06 - 1.24	9.18e-04	1.16	1.06 - 1.28	1.26e-03
Pt < 0.005	1.21	1.12 - 1.31	1.25e-06	1.25	1.14 - 1.37	1.90e-06
Pt < 0.01	1.22	1.13 - 1.32	3.90e-07	1.25	1.14 - 1.38	1.31e-06
Pt < 0.05	1.23	1.14 - 1.33	2.27e-07	1.27	1.16 - 1.40	2.81e-07
Pt < 0.1	1.25	1.16 - 1.36	1.45e-08	1.31	1.19 - 1.44	1.57e-08
Pt < 0.5	1.23	1.14 - 1.34	1.15e-07	1.27	1.15 - 1.39	6.35e-07
Pt < 1.0	1.22	1.13 - 1.32	3.26e-07	1.25	1.14 - 1.37	1.66e-06
POOLED	Cases 3141, Controls 3182			Cases: 1819, Controls: 3182		
Pt < 0.0001	1.13	1.08 - 1.20	4.14e-06	1.17	1.10 - 1.24	1.55e-06
Pt < 0.001	1.19	1.12 - 1.25	2.94e-10	1.23	1.16 - 1.31	6.96e-11
Pt < 0.005	1.25	1.19 - 1.32	1.89e-16	1.30	1.22 - 1.39	1.07e-16
Pt < 0.01	1.25	1.19 - 1.32	6.48e-17	1.31	1.23 - 1.39	3.72e-17
Pt < 0.05	1.26	1.19 - 1.33	3.62e-17	1.31	1.23 - 1.40	9.81e-18
Pt < 0.1	1.28	1.22 - 1.35	7.95e-20	1.33	1.25 - 1.42	2.03e-19
Pt < 0.5	1.28	1.21 - 1.35	4.13e-19	1.32	1.24 - 1.40	6.91e-18
Pt < 1.0	1.27	1.20 - 1.34	2.54e-18	1.31	1.23 - 1.39	4.53e-17

URL

Summary statistics
PGC: http://www.med.unc.edu/pgc/downloads

Software
PLINK: http://pngu.mgh.harvard.edu/~purcell/plink/

GCTA: http://www.complextraitgenomics.com/software/gcta/

AVENGEME: https://sites.google.com/site/fdudbridge/software/ version 2015-10-12

LD-PRED: https://github.com/bvilhjal/ldpred

SUPPLEMENTAL REFERENCES

1. Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, *et al.* (2008): Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet.* 16: 335–342.

2. Peyrot WJ, Milaneschi Y, Abdellaoui A, Sullivan PF, Hottenga JJ, Boomsma DI, Penninx BWJH (2014): Effect of polygenic risk scores on depression in childhood trauma. *Br J Psychiatry.* 205: 113–9.

3. Abdellaoui A, Hottenga J-J, de Knijff P, Nivard MG, Xiao X, Scheet P, *et al.* (2013): Population structure, migration, and diversifying selection in the Netherlands. *Eur J Hum Genet.* 21: 1277–85.

4. Milaneschi Y, Lamers F, Peyrot WJ, Abdellaoui A, Willemsen G, Hottenga J, *et al.* (2016): Polygenic dissection of major depression clinical heterogeneity. *Mol Psychiatry.* 21: 516–22.

5. Yang J, Lee SH, Goddard ME, Visscher PM (2011): GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet.* 88: 76–82.

6. Major Depressive Disorder working groups of the Psychiatric GWAS Consortium, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, *et al.* (2013): A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry.* 18: 497–511.

7. Sklar P, Ripke S, Scott LJ, Andreassen O a, Cichon S, Craddock N, *et al.* (2011): Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet.* 43: 977–983.

8. Schizophrenia Working Group of the Psychiatric Genomics (2014): Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 511: 421–7.

9. Sullivan PF, de Geus EJC, Willemsen G, James MR, Smit JH, Zandbelt T, *et al.* (2009): Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry.* 14: 359–75.

10. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al.* (2007): PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81: 559–75.

11. Vilhjálmsson BJ, Yang J, Finucane H (2015): Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *bioRxiv.* .

12. Peyrot WJ, Lee SH, Milaneschi Y, Abdellaoui A, Byrne EM, Esko T, *et al.* (2015): The association between lower educational attainment and depression owing to shared genetic effects? Results in ~25,000 subjects. *Mol Psychiatry.* 20: 735–43.

13. Fava GA, Kellner R (1993): Staging: a neglected dimension in psychiatric classification. *Acta Psychiatr Scand.* .

14. Mcgorry PD, Hickie IB, Yung AR, Pantelis C, Jackson HJ (2006): Clinical staging of psychiatric disorders: a heuristic framework for choosing earlier, safer and more effective interventions. *Aust N Z J Psychiatry.* 616–622.

15. Hetrick SE, Parker a G, Hickie IB, Purcell R, Yung a R, McGorry PD (2008): Early identification and intervention in depressive disorders: towards a clinical staging model. *Psychother Psychosom.* 77: 263–70.

16. Hermens DF, Naismith SL, Lagopoulos J, Lee RSC, Guastella AJ, Scott EM, Hickie IB (2013): Neuropsychological profile according to the clinical stage of young persons presenting for mental health care. *BMC Psychol.* 1: 8.

17. Lagopoulos J, Hermens DF, Hatton SN, Battisti R a, Tobias-Webb J, White D, *et al.* (2013): Microstructural white matter changes are correlated with the stage of psychiatric illness. *Transl Psychiatry.* 3: e248.

18. Lagopoulos J, Hermens DF, Naismith SL, Scott EM, Hickie IB (2012): Frontal lobe changes occur early in the course of affective disorders in young people. *BMC Psychiatry.* 12: 4.

19. Verduijn J, Milaneschi Y, van Hemert AM, Schoevers RA, Hickie IB, Penninx BWJH, Beekman ATF (2015): Clinical Staging of Major Depressive Disorder: An Emperical Exploration. *J Clin Psychiatry.* 76: 1200–1208.

20. Verduijn J, Milaneschi Y, Schoevers RA, van Hemert AM, Beekman ATF, Penninx BWJH (2015): Pathophysiology of Major Depressive Disorder: mechanisms involved in etiology are not associated with clinical progression. *Transl Psychiatry.* 5: 1–9.

21. Hickie IB, Scott EM, Hermens DF, Naismith SL, Guastella AJ, Kaur M, *et al.* (2013): Applying clinical staging to young people who present for mental health care. *Early Interv Psychiatry.* 7: 31–43.

22. Kapczinski F, Dias VV, Kauer-Sant'Anna M, Brietzke E, Vázquez GH, Vieta E, Berk M (2009): The potential use of biomarkers as an adjunctive tool for staging bipolar disorder. *Prog Neuropsychopharmacol Biol Psychiatry.* 33: 1366–71.

23. Lee SH, Ripke S, Neale BM, Faraone S V, Purcell SM, Perlis RH, *et al.* (2013): Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet.* 45: 984–94.

24. Sklar P, Ripke S, Scott L, Andreassen O (2011): Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet.* 1–37.

25. Dudbridge F (2013): Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 9: e1003348.